

Figure 1

Num	PDB Code	Norm Align Score	Raw Align Score	%ID	%Struct Aligned	%Query Seq Aligned	Pairwise Energy	Solvation Energy	Neural Net Score	%Confidence	From Query	To Pos for Query	From Pos for Target	To Pos for Target	Alignment Length	Local-Q	Global-Q
1	143MBB biopendium (align)	38.21	60	37.8	31.5	46.2	-17.99	-0.17	0.957	84	30	66	20	55	37		0
2	143MBB biopendium (align)	36.91	59	37.8	29.4	46.2	-18.54	0.08	0.953	83	30	66	28	63	37		0
3	143CABD biopendium (align)	36.84	59	37.8	28.9	46.2	-17.34	1.14	0.951	82	30	66	30	65	37		0
4	143CABD biopendium (align)	33.91	56	37.8	29.4	46.2	-8.35	0.60	0.940	78	30	66	30	65	37		0
5	143CABD biopendium (align)	30.16	56	31.9	15.5	59.0	-15.63	-0.64	0.922	72	21	67	190	234	47		0

<<Previous 1 Next>>

Figure 2

Alignment

```

Alignment Type: Local
Sequence A Range: 1 -> 111
Sequence B Range: 1 -> 78
Gap Open Penalty: -11
Gap Extend Penalty: -1
Scoring Matrix: /usr/local/BLOSUM62
Profile A: ../gtws_files/profiles/ld9gBB00.pro
Sequence B: /tmp/gtw_6314.fa
DB Alignment: -
GT Alignment: -
View Alignment: Yes
Reverse GT Alignment: No
Score Length Num_ID No.+va OvrIp %ID %+ve From To From To
SCORES: 60 37 14 19 68 37.8 51.4 20 55 30 66
Length1 Length2 Normalised-Score
SCORE2: 111 78 38.210598

10| 20| 30| 40| 50| 60
ld9gBB00 -----qffreienlkeyfnggplfSEILKNWKDESDKKIIQSQIVS-FYFKLPENLKDNQViqrs
IPAAA445 mtspnelnkplwtnpgeteicdlsdtefkISVLKNLKEIQDNTEKESRILSDKYKKQIEIIKGNQaeile
10| 20| 30| 40| 50| 60| 70|

70| 80| 90| 100| 110|
ld9gBB00 mdiikqdmfqkflngssekledfkkliqipvddlqiqrkainelikvmndls
IPAAA445 lrnadgtl-----

```

Figure 3

INSP037 (IPAAA44548) Predicted sequence with translation product:

```

1  TGCCTAGACA CCAAAGAACA ACTATTACCA TCAACAACAT CCAGTAAAC ATGACTTCAC CAAACGAACT
                                     m t s p n e

71  AAATAAGCTG CCATGGACCA ATCCTGGAGA AACAGAGATA TGTGACCTTT CAGACACAGA ATTCAAATA
    l n k l p w t n p g e t e i c d l s d t e f k i

141 TCTGTGTTGA AGAACCTCAA AGAAATTCAA GATAACACAG AGAAGGAATC CAGAATTCTA TCAGACAAAT
    s v l k n l k e i q d n t e k e s r i l s d k

211 ATAAGAAACA GATTGAAATA ATTAAAGGGA ATCAAGCAGA AATTCTGGAG TTGAGAAATG CAGATGGCAG
    y k k q i e i i k g n q a e i l e l r n a d g

281 ACTTTAGAAT GCATAAGAGT CTTTTATAG CAGAATTCAT CAAGCAGAAG AAAGAAT
    t l

```

The position of primers is denoted by the shaded boxes above.

Figure 4**INSP037 (IPAAA44548) Cloned sequence with translation**

```
1  GCATCAACAA CATCCAGTAA AACATGACTT CACCAAACGA ACTAAATAAG CTGCCATGGA CCAATCCTGG
    m t s p n e l n k l p w t n p

71  AGAAACAGAG ATATGTGACC TTTCAGACAC AGAATTCAAA ATATCTGTGT TGAAGAACCT CAAGGAAATT
    g e t e i c d l s d t e f k i s v l k n l k e i

141 CAAGATAACA CAGAGAAGGA ATCCAGAATT CTATCAGACA AATATAAGAA ACAGATTGAA ATAATTAAAG
    q d n t e k e s r i l s d k y k k q i e i i k

211 GGAATCAAGC AGAAATTCTG GAGTTGAGAA ATGCAGATGG CACACTTTAG AATG
    g n q a e i l e l r n a d g t l
```

Figure 5

Map of PCRII-TOPO-IPAAA44548

Molecule: pCRII-TOPO-IPAAA44548, 4214 bps DNA Circular
 File Name: 13124.cm5
 Description: Plasmid ID 13124

Molecule Features:

Type	Start	End	Name	Description
MARKER	239		SP6	
REGION	337	600		IPAAA44548 cloned sequence
GENE	577	341	C 44548 cds	
MARKER	670		C T7	
REGION	854	1268	f1 ori	
GENE	1602	2396	KanR	
GENE	2414	3274	AmpR	
REGION	3419	4092	pUC ori	

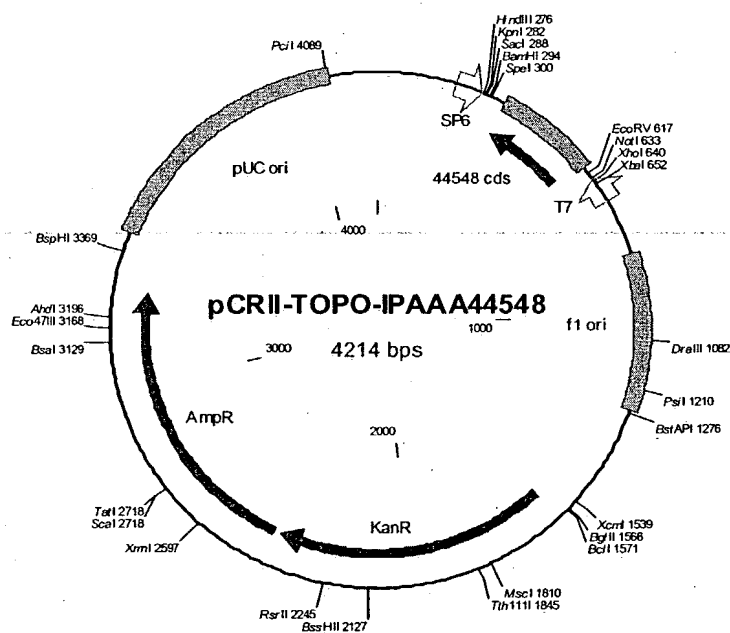


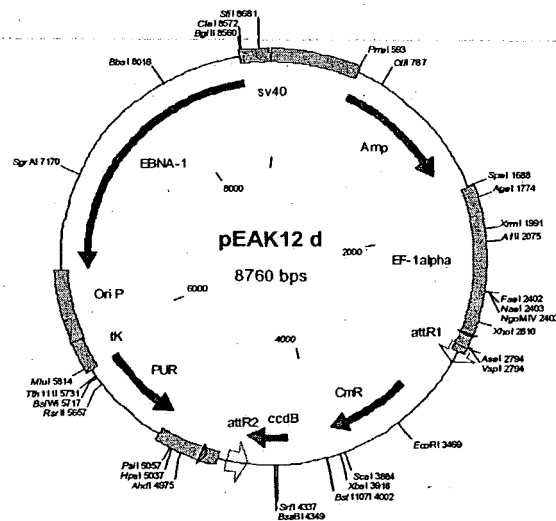
Figure 6

Map of expression vector pEAK12d

Molecule: pEAK12 d, 8760 bps DNA Circular
File Name: pEAK12DEST.cm5
Description: Mammalian cell expression vector (plasmid ID 11345)

Molecule Features:

Type	Start	End	Name	Description
REGION	2	595		pmb-ori
GENE	596	1519	Amp	
REGION	1690	2795	EF-1alpha	
REGION	2703	2722		position of pEAK12F primer
REGION	2796	2845		MCS
MARKER	2855		attR1	
GENE	3256	3915	CmR	
GENE	4257	4562	ccdB	
MARKER	4603		C attR2	
REGION	4733	4733		MCS
REGION	4734	5162		poly A/splice
REGION	4819	4848	C	position of pEAK12R primer
GENE	5781	5163	C PUR	PUROMYCIN
REGION	6005	5782	C tK	tK promoter
REGION	6500	6006	C Ori P	
GENE	8552	6500	C EBNA-1	
REGION	8553	8752	sv40	



7/24

Figure 7

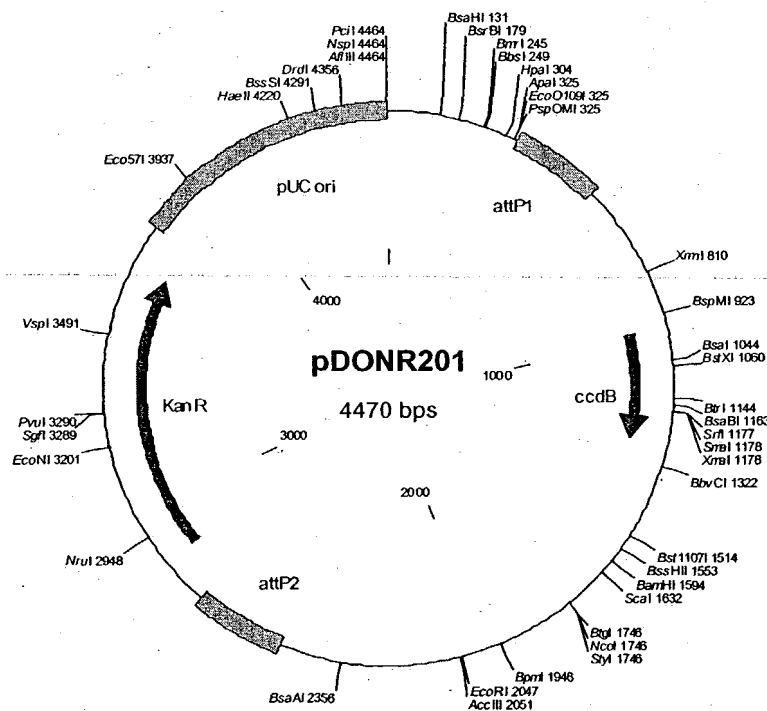
Map of plasmid pDONR201

Molecule: pDONR201, 4470 bps DNA Circular
File Name: pDONR201.cm5, dated 17 Oct 2002

Description: Gateway entry vector (Invitrogen)- plasmid ID# 13309

Molecule Features:

Type	Start	End	Name
REGION	332	563	attP1
GENE	959	1264	ccdB
REGION	2513	2744	attP2
GENE	2868	3677	KanR
REGION	3794	4467	pUC ori



8/24

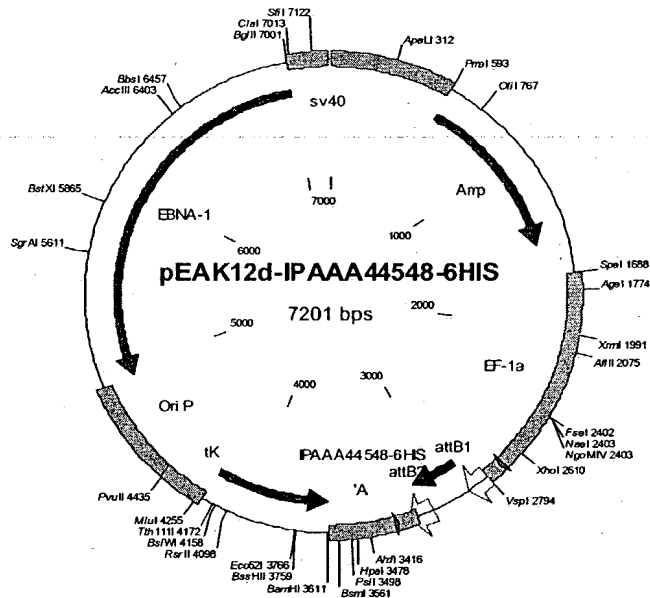
Figure 8

Map of expression vector pEAK12d-IPAAA44548-6HIS

Molecule: pEAK12d-IPAAA44548-6HIS, 7201 bps DNA Circular
 File Name: 11775.cm5
 Description: Mammalian cell Expression Construct

Molecule Features:

Type	Start	End	Name	Description
REGION	2	595		pmb-ori
GENE	596	1519	Amp	
REGION	1690	2795	EF-1a	
REGION	2703	2722		peak12D-F primer
MARKER	2855		attB1	
GENE	2888	3139	IPAAA44548-6HIS	
MARKER	3155		attB2	
REGION	3175	3603	'A	poly A/splice
REGION	3289	3270	C	pEAK12D-R primer
GENE	4222	3604	C	PUROMYCIN
REGION	4446	4223	C tK	tK promoter
REGION	4941	4447	C Ori P	
GENE	6993	4941	C EBNA-1	
REGION	6994	7193	sv40	



9/24

Figure 9

Map of E.coli expression vector pDEST14

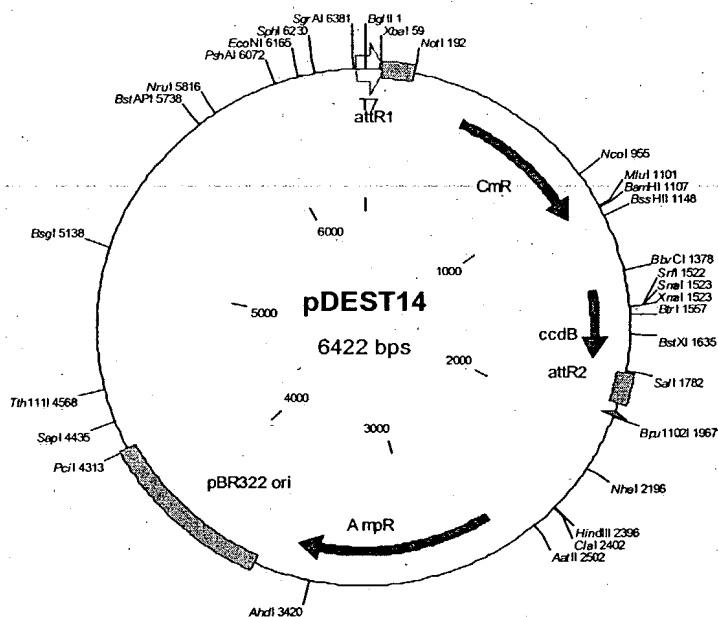
Molecule: pDEST14, 6422 bps DNA Circular
File Name: pDEST14.cm5, dated 17 Oct 2002

Description: E.coli expression vector (Invitrogen)

Notes: Gateway compatible, Expression under control of T7 promoter

Molecule Features:

Type	Start	End	Name	Description
MARKER	21		T7	Promoter
REGION	67	191	attR1	
GENE	441	1100	CmR	
GENE	1442	1747	ccdB	
REGION	1788	1912	attR2	
REGION	1964	1944 C		pDEST14 R primer
GENE	2638	3498	AmpR	
REGION	3643	4316	pBR322 ori	



10/24

Figure 10

Map of plasmid pDEST14-IPAAA44548-6HIS

Molecule: pDEST14-IPAAA44548-6HIS, 4899 bps DNA Circular
 File Name: 12896.cm5
 Description: plasmid ID 12896

Molecule Features:

Type	Start	End	Name	Description
MARKER	21		T7	
REGION	72	67	C attB1	
REGION	94	108		Shine Dalgarno Sequence
GENE	109	360	IPAAA44548-6HIS	
REGION	376	389	attB2	
REGION	441	421	C	pDEST14-R primer
GENE	1115	1975	Amp	
REGION	2124	2763	ori	pBR322 ori

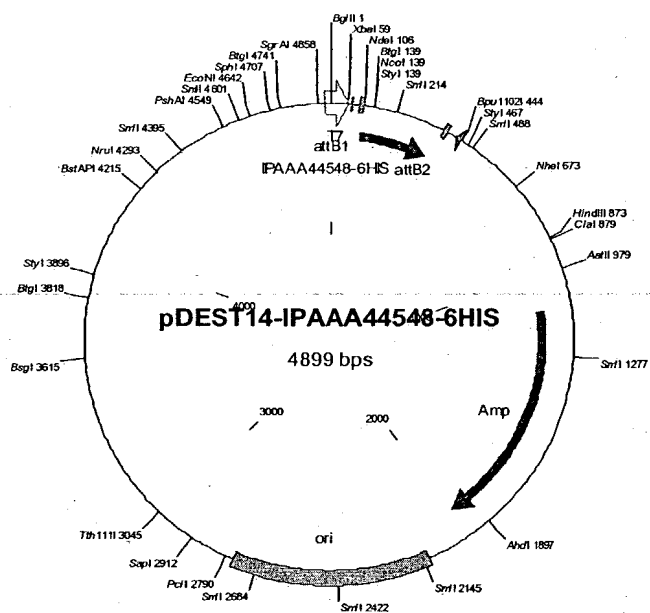


Figure 11

PCRII TOPO IPAAA44548

1 AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC
 61 ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC
 121 TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA
 181 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT
 241 TTAGGTGACA CTATAGAATA CTCAAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA
 301 CTAGTAACGG CCGCCAGTGT GCTGGAATTC GCCCTTCATT CTAAAGTGTG CCATCTGCAT
 361 TTCTCAACTC CAGAATTTCT GCTTGATTCC CTTTAATTAT TTCAATCTGT TTCTTATATT
 421 TGTCTGATAG AATTCTGGAT TCCTTCTCTG TGTATCTTG AATTTCTTG AGGTTCTTCA
 481 ACACAGATAT TTTGAATTCT GTGTCTGAAA GGTCACATAT CTCTGTTTCT CCAGGATTGG
 541 TCCATGGCAG CTTATTTAGT TCGTTTGGTG AAGTCATGTT TTACTGGATG TTGTTGATGC
 601 AAGGGCGAAT TCTGCAGATA TCCATCACAC TGGCGGCCGC TCGAGCATGC ATCTAGAGGG
 661 CCCAATTGCG CCTATAGTGA GTCGTATTAC AATTCAGTGG CCGTCGTTTT ACAACGTCGT
 721 GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG CAGCACATCC CCCTTTCGCC
 781 AGCTGGCGTA ATAGCGAAGA GGCCCGCACC GATCGCCCTT CCCAACAGTT GCGCAGCCTG
 841 AATGGCGAAT GGGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG
 901 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT
 961 TCCTTTCTCG CCACGTTGCG CGGCTTTCCC CGTCAAGCTC TAAATCGGGG GCTCCCTTTA
 1021 GGGTTCGAT TTAGAGCTTT ACGGCACCTC GACCGCAAAA AACTTGATTT GGGTGATGGT
 1081 TCACGTAGTG GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG
 1141 TTCTTTAATA GTGGACTCTT GTTCCAACT GGAACAACAC TCAACCCTAT CGCGGTCTAT
 1201 TCTTTTGATT TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAA TGAGCTGATT
 1261 TAACAAATTC AGGGCGCAAG GGCTGCTAAA GGAACCGGAA CACGTAGAAA GCCAGTCCGC
 1321 AGAAACGGTG CTGACCCCGG ATGAATGTCA GCTACTGGGC TATCTGGACA AGGGAAAACG
 1381 CAAGCGCAA GAGAAAGCAG GTAGCTTGCA GTGGGCTTAC ATGGCGATAG CTAGACTGGG
 1441 CGGTTTTATG GACAGCAAGC GAACCGGAAT TGCCAGCTGG GGCGCCCTCT GGTAAGGTTG

1501 GGAAGCCCTG CAAAGTAAAC TGGATGGCTT TCTTGCCGCC AAGGATCTGA TGGCGCAGGG
1561 GATCAAGATC TGATCAAGAG ACAGGATGAG GATCGTTTCG CATGATTGAA CAAGATGGAT
1621 TGCACGCAGG TTCTCCGGCC GCTTGGGTGG AGAGGCTATT CGGCTATGAC TGGGCACAAC
1681 AGACAATCGG CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC AGCGCAGGGG CGCCCGGTTC
1741 TTTTGTCAA GACCGACCTG TCCGGTGCCC TGAATGAACT GCAGGACGAG GCAGCGCGGC
1801 TATCGTGGCT GGCCACGACG GCGGTTCCCTT GCGCAGCTGT GCTCGACGTT GTCACTGAAG
1861 CGGGAAGGGA CTGGCTGCTA TTGGGCGAAG TGCCGGGGCA GGATCTCCTG TCATCTCGCC
1921 TTGCTCCTGC CGAGAAAGTA TCCATCATGG CTGATGCAAT GCGGCGGCTG CATACGCTTG
1981 ATCCGGCTAC CTGCCCATTC GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTACTC
2041 GGATGGAAGC CGGTCTTGTC GATCAGGATG ATCTGGACGA AGAGCATCAG GGGCTCGCGC
2101 CAGCCGAACT GTTCGCCAGG CTCAAGGCGC GCATGCCCCA CGGCGAGGAT CTCGTCGTGA
2161 TCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA TGGCCGCTTT TCTGGATTCA
2221 ACCACTGTGG CCGGCTGGGT GTGGCGGACC GCTATCAGGA CATAGCGTTG GATACCCGTG
2281 ATATTGCTGA AGAGCTTGGC GGCGAATGGG CTGACCGCTT CCTCGTGCTT TACGGTATCG
2341 CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAATTG
2401 AAAAAGGAAG AGTATGAGTA TTCAACATTT CCGTGTGCGC CTTATTCCCT TTTTGCGGC
2461 ATTTTGCCTT CCTGTTTTTG CTCACCCAGA AACGCTGGTG AAAGTAAAAG ATGCTGAAGA
2521 TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA AGATCCTTGA
2581 GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTAAAGTTC TGCTATGTGA
2641 TACACTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTGCGCGCA TACACTATTC
2701 TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC
2761 AGTAAGAGAA TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT
2821 TCTGACAACG ATCGGAGGAC CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA
2881 TGTAACTCGC CTTGATCGTT GGAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGAG
2941 TGACACCACG ATGCCTGTAG CAATGCCAAC AACGTTGCGC AAACATTAA CTGGCGAACT
3001 ACTTACTCTA GCTTCCCGGC AACAAATTAAT AGACTGAATG GAGGCGGATA AAGTTGCAGG
3061 ACCACTTCTG CGCTCGGCCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG

3121 TGAGCGTGGG TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC GCTCCCGTAT
3181 CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC
3241 TGAGATAGGT GCCTCACTGA TTAAGCATTG GTAAGTGTCA GACCAAGTTT ACTCATATAT
3301 ACTTTAGATT GATTTAAAC TTCATTTTTA ATTTAAAGG ATCTAGGTGA AGATCCTTTT
3361 TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTTCG TTCCACTGAG CGTCAGACCC
3421 CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT
3481 GCAAACAAAA AAACCACCGC TACCAGCGGT GGTGTGTTG CCGGATCAAG AGCTACCAAC
3541 TCTTTTTCCG AAGGTAAGT GCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT
3601 GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT
3661 GCTAATCCTG TTACCAGTGG CTGCTGCCAG TGGCGATAAG TCGTGTCTTA CCGGGTTGGA
3721 CTCAAGACGA TAGTTACCGG ATAAGGCGCA GCGGTCGGGC TGAACGGGGG GTTCGTGCAC
3781 ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA TACCTACAGC GTGAGCTATG
3841 AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA GCGGCAGGGT
3901 CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAAC GCCTGGTATC TTTATAGTCC
3961 TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTGT TGATGCTCGT CAGGGGGGCG
4021 GAGCCTATGG AAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGGCT TTTGCTGGCC
4081 TTTTGCTCAC ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC
4141 CTTTGAGTGA GCTGATACCG CTCGCCGAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG
4201 CGAGGAAGCG GAAG

Figure 12

pDEST14-IPAAA44548-6HIS

1 AGATCTCGAT CCCGCGAAAT TAATACGACT CACTATAGGG AGACCACAAC GGTTCCTC TAGATCACAA GTTTGTACAA
 81 AAAAGCAGGC TTCGAAGGAG ATATACATAT GACTTCACCA AACGAACTAA ATAAGCTGCC ATGGACCAAT CCTGGAGAAA
 161 CAGAGATATG TGACCTTTCA GACACAGAAT TCAAATATC TGTGTTGAAG AACCTCAAGG AAATTCAAGA TAACACAGAG
 241 AAGGAATCCA GAATTCTATC AGACAAATAT AAGAAACAGA TTGAAATAAT TAAAGGGAAT CAAGCAGAAA TTCTGGAGTT
 321 GAGAAATGCA GATGGCACAC TTCACCATCA CCATCACCAT TGAAACCCAG CTTTCTTGTA CAAAGTGGTG ATGATCCGGC
 401 TGCTAACAAA GCCCGAAAGG AAGCTGAGTT GGCTGCTGCC ACCGCTGAGC AATAACTAGC ATAACCCCTT GGGCCTCTA
 481 AACGGGTCTT GAGGGGTTTT TTGCTGAAAG GAGGAACTAT ATCCGGATAT CCACAGGACG GGTGTGGTCG CCATGATCGC
 561 GTAGTCGATA GTGGCTCCAA GTAGCGAAGC GAGCAGGACT GGGCGGCGGC CAAACCGGTC GGACAGTGCT CCGAGAACGG
 641 GTGCGCATAG AAATTGCATC AACGCATATA GCGCTAGCAG CACGCCATAG TGAATGGCGA TGCTGTCCGA ATGGACGATA
 721 TCCCGCAAGA GGCCCGGCAG TACCGGCATA ACCAAGCCTA TGCTACAGC ATCCAGGGTG ACGGTGCCGA GGATGACGAT
 801 GAGCGCATTG TTAGATTTCA TACACGGTGC CTGACTGCGT TAGCAATTTA ACTGTGATAA ACTACCGCAT TAAAGCTTAT
 881 CGATGATAAG CTGTCAAACA TGAGAATTCT TGAAGACGAA AGGGCCTCGT GATACGCCTA TTTTATAGG TTAATGTCAT
 961 GATAATAATG GTTCTTAGA CGTCAGGTGG CACTTTTCGG GGAAATGTGC GCGGAACCCC TATTTGTTTA TTTTCTAAA
 1041 TACATTCAAA TATGTATCCG CTCATGAGAC AATAACCCTG ATAAATGCTT CAATAATATT GAAAAAGGAA GAGTATGAGT
 1121 ATTCAACATT TCCGTGTCGC CCTTATTCCT TTTTTCGGC CATTTTGCTT TCCTGTTTTT GCTCACCAG AAACGCTGGT
 1201 GAAAGTAAAA GATGCTGAAG ATCAGTTGGG TGCACGAGTG GGTACATCG AACTGGATCT CAACAGCGGT AAGATCCTTG
 1281 AGAGTTTTCG CCCGAAGAA CGTTTCCAA TGATGAGCAC TTTTAAAGTT CTGCTATGTG GCGCGGTATT ATCCCGTGT
 1361 GACGCCGGC AAGAGCAACT CGGTGCGCGC ATACACTATT CTCAGAATGA CTTGGTTGAG TACTCACCAG TCACAGAAAA
 1441 GCATCTTACG GATGGCATGA CAGTAAGAGA ATTATGCAGT GCTGCCATAA CCATGAGTGA TAACACTGCG GCCAACTTAC
 1521 TTCTGACAAC GATCGGAGGA CCGAAGGAGC TAACCGCTT TTTGACAAC ATGGGGGATC ATGTAAGTCG CCTTGATCGT
 1601 TGGGAACCGG AGCTGAATGA AGCCATACCA AACGACGAGC GTGACACCAC GATGCCTGCA GCAATGGCAA CAACGTTGCG
 1681 CAAACTATTA ACTGGCGAAC TACTTACTCT AGCTTCCCGG CAACAATTAA TAGACTGGAT GGAGGCGGAT AAAGTTGCAG
 1761 GACCACTTCT GCGCTCGGCC CTTCGGGTG GCTGGTTTAT TGCTGATAAA TCTGGAGCCG GTGAGCGTGG GTCTCGCGGT
 1841 ATCATTGCAG CACTGGGGCC AGATGGTAAG CCCTCCCGTA TCGTAGTTAT CTACACGAGC GGGAGTCAGG CAACTATGGA
 1921 TGAACGAAAT AGACAGATCG CTGAGATAGG TGCCTCACTG ATTAAGCATT GGTAAGTGC AGACCAAGTT TACTCATATA
 2001 TACTTTAGAT TGATTTAAAA CTTTATTTT AATTTAAAG GATCTAGGTG AAGATCCCTT TTGATAATCT CATGACCAAA
 2081 ATCCCTTAAC GTGAGTTTTC GTTCCACTGA GCGTCAGACC CCGTAGAAAA GATCAAAGGA TCTTCTGAG ATCCTTTTTT
 2161 TCTGCGCGTA ATCTGCTGCT TGCAACAAA AAAACCACCG CTACCAGCGG TGGTTGTTT GCCGGATCAA GAGCTACCAA
 2241 CTCTTTTCC GAAGGTAAGT GGCTTCAGCA GAGCGCAGAT ACCAAATACT GTCCTTCTAG TGTAGCCGTA GTTAGGCCAC
 2321 CACTTCAAGA ACTCTGTAGC ACCGCCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG GCTGCTGCCA GTGGCGATAA

2401 GTCGTGTCTT ACCGGGTTGG ACTCAAGACG ATAGTTACCG GATAAGGCGC AGCGGTCGGG CTGAACGGGG GGTTCGTGCA
 2481 CACAGCCCAG CTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG CGTGAGCTAT GAGAAAGCGC CACGCTTCCC
 2561 GAAGGGAGAA AGCGGACAG GTATCCGTA AGCGGCAGG TCGAACAGG AGAGCGCAG AGGGAGCTTC CAGGGGAAA
 2641 CGCCTGGTAT CTTTATAGTC CTGTCGGTT TCGCCACCTC TGACTTGAGC GTCGATTTT GTGATGCTCG TCAGGGGGC
 2721 GGAGCCTATG GAAAAACGCC AGCAACGCG CCTTTTACG GTTCCTGGCC TTTTGCTGCC CTTTGTCTCA CATGTTCTTT
 2801 CCTGCGTTAT CCCCTGATTC TGTGGATAAC CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGA GCCGAACGAC
 2881 CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC GGAAGAGCG CTGATGCGGT ATTTTCTCCT TACGCATCTG TCGGTATTT
 2961 CACACCGCAT ATATGGTGCA CTCTCAGTAC AATCTGCTCT GATGCCGCAT AGTTAAGCCA GTATACACTC CGCTATCGCT
 3041 ACGTGA CTGG GTCATGGCTG CGCCCCGACA CCCGCCAACA CCCGCTGACG CGCCCTGACG GGCTTGTCTG CTCCCGCAT
 3121 CCGCTTACAG ACAAGCTGTG ACCGTCTCCG GGAGCTGCAT GTGTCAGAGG TTTTCACCGT CATCACCGAA ACGCGCGAGG
 3201 CAGCTGCGGT AAAGCTCATC AGCGTGGTCG TGAAGCGATT CACAGATGTC TGCTGTTC TCCGCTCCA GCTCGTGGAG
 3281 TTTCTCCAGA AGCGTTAATG TCTGGCTTCT GATAAAGCGG GCCATGTTAA GGGCGGTTT TTCTGTTTG GTCAGTATG
 3361 CCTCCGTGTA AGGGGATTT CTGTTATGG GGGTAATGAT ACCGATGAAA CGAGAGAGGA TGCTCACGAT ACGGGTACT
 3441 GATGATGAAC ATGCCCGGTT ACTGGAACGT TGTGAGGGTA AACAACTGGC GGTATGGATG CGGCGGGACC AGAGAAAAAT
 3521 CACTCAGGGT CAATGCCAGC GCTTCGTTAA TACAGATGTA GGTGTTCCAC AGGGTAGCCA GCAGCATCCT GCGATGCAGA
 3601 TCCGGAACAT AATGGTGAG GCGCTGACT TCCGCGTTT CAGACTTTAC GAAACACGGA AACCAGAAC CATTCATGTT
 3681 GTTGCTCAGG TCGCAGACGT TTTGACAGC CAGTCGCTTC ACGTTCGCTC GCGTATCGGT GATTCACTCT GCTAACAGT
 3761 AAGGCAACCC CGCCAGCCTA GCCGGTCTT CAACGACAGG AGCACGATCA TCGCACCCG TGGCCAGGAC CCAACGCTGC
 3841 CCGAGATGCG CGCGTGCGG CTGCTGGAGA TGGCGGACGC GATGGATATG TTCTGCCAAG GGTGTTGTTG CGCATTCACA
 3921 GTTCTCCGA AGAATTGATT GGCTCCAATT CTTGGAGTGG TGAATCCGTT AGCGAGGTGC CGCCGGCTTC CATTCAGGTC
 4001 GAGGTGGCCC GGCTCCATGC ACCGCGACGC AACCGGGGA GGCAGACAAG GTATAGGGCG GCGCTACAA TCCATGCCAA
 4081 CCCGTTCCAT GTGCTCGCG AGGCGGCATA AATCGCCGTG ACGATCAGCG GTCCAGTGAT CGAAGTTAGG CTGGTAAGAG
 4161 CCGCGAGCGA TCCTTGAAGC TGTCCCTGAT GGTGTCATC TACCTGCCTG GACAGCATGG CCTGCAACGC GGGCATCCCG
 4241 ATGCCGCCCG AAGCGAGAAG AATCATAATG GGAAGGCCA TCCAGCCTCG CGTCGCGAAC GCCAGCAAGA CGTAGCCAG
 4321 CGCGTCGGC GCCATGCCG CGATAATGGC CTGCTTCTCG CCGAAACGTT TGGTGGCGG ACCAGTGACG AAGGCTGAG
 4401 CGAGGGCGTG CAAGATTCCG AATACCGCAA GCGACAGGCC GATCATCGTC GCGCTCCAGC GAAAGCGGTC CTCGCCGAAA
 4481 ATGACCCAGA GCGCTGCCG CACCTGTCTT ACGAGTTGCA TGATAAAGAA GACAGTCATA AGTGGCGGA CGATAGTCAT
 4561 GCCCCGCGC CACCGGAAG AGCTGACTGG GTTGAAGGCT CTCAAGGGCA TCGGTCGATC GACGCTCTCC CTTATGCGAC
 4641 TCCTGCATTA GGAAGCAGC CAGTAGTAGG TTGAGGCCG TGAGACCGC CGCCGCAAGG AATGGTGCAT GCAAGGAGAT
 4721 GCGGCCAAC AGTCCCCCG CCACGGGGC TGCCACCATA CCCACCGCA AACAAAGCGCT CATGAGCCCG AAGTGGCGAG
 4801 CCCGATCTT CCCATCGGTG ATGTCGGCGA TATAGGCGC AGCAACCGCA CCTGTGGGCG CGGTGATGCC GGCCACGATG
 4881 CGTCCGGCGT AGAGGATCG

Figure 13

pEAK12D-IPAAA44548-6HIS

1 GGCCTAATCT GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC TACCAACTCT
 81 TTTTCCGAAG GTAAGTGGCT TCAGCAGAGC GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCCACCACT
 161 TCAAGAACTC TGTAGCACCG CCTACATACC TCGCTCTGCT GAAGCCAGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC
 241 GTGTCTTACC GGGTTGGACT CAAGAGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT CGTGACACA
 321 GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG CTTCCCGAAG
 401 GGAGAAAGGC GGACAGGTAT CCGTAAGCG GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC
 481 TGGTATCTTT ATAGTCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTGTGA TGCTCGTCAG GGGGGCGGAG
 561 CCTATGAAA AACGCCAGCA ACGCAAGCTA GAGTTTAAAC TTGACAGATG AGACAATAAC CCTGATAAAT GCTTCAATAA
 641 TATTGAAAA GGAAGATAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT
 721 TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCA GAAGATCACT TGGGTGCGCG AGTGGGTAC ATCGAACTGG
 801 ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTC CCAATGATGA GCACTTTTAA AGTTCTGCTA
 881 TGTGGCGCGG TATTATCCCG TATTGATGCC GGGCAAGAGC AACTCGGTGC CCGCATACAC TATTCTCAGA ATGACTTGGT
 961 TGAATACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA
 1041 GTGATAACAC TGCGGCCAAC TTAATTCTGA CAACTATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG
 1121 GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC
 1201 TGTAGCAATG GCAACAACGT TCGGAAAACCT ATTAAGTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA CTAATAGACT
 1281 GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC GGCACCTCCG GCTGGCTGGT TTATTGCTGA TAAATCAGGA
 1361 GCCGGTGAGC GTGGGTCACG CCGTATCATT GCAGCACTGG GGCCGGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC
 1441 TACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAG
 1521 GATAAATTTC TGGTAAGGAG GACACGTATG GAAGTGGGCA AGTTGGGGAA GCCGTATCCG TTGCTGAATC TGGCATATGT
 1601 GGGAGTATAA GACGCGCAGC GTCGCATCAG GCATTTTTTT CTGCGCCAAT GCAAAAAGGC CATCCGTCAG GATGGCCTTT
 1681 CGGCATAACT AGTGAGGCTC CCGTGCCCGT CAGTGGGCAG AGCGCACATC GCCCACAGTC CCCGAGAAGT TGGGGGGAGG
 1761 GGTGCGCAAT TGAACCGGTG CCTAGAGAAG GTGGCGCGGG GTAAACTGGG AAAGTGATGT CGTGTACTGG CTCGCCCTTT
 1841 TTCCCGAGGG TGGGGGAGAA CCGTATATAA GTGCAGTAGT CGCCGTGAAC GTTCTTTTTT GCAACGGGTT TGCCGCCAGA
 1921 ACACAGGTAA GTGCCGTGTG TGGTTCCGCG GGGCTTGGCC TCTTTACGGG TTATGGCCCT TCGCTGCCCT GAATTACTTC
 2001 CACCTGGCTG CAGTACGTGA TTCTTGATCC CGAGCTTCGG GTTGGAAAGT GGTGGGAGAG TTCGAGGCCT TGCGCTTAAG
 2081 GAGCCCTTC GCCTCGTGCT TGAGTTGAGG CCTGGCCTGG GCGCTGGGGC CGCCGCGTGC GAATCTGGTG GCACCTTCGC
 2161 GCCTGTCTCG CTGCTTTCGA TAAGTCTCTA GCCATTAAAA ATTTTGTATG ACCTGCTGCG ACGCTTTTTT TCTGGCAAGA

2241 TAGTCTTGTA AATGCGGGCC AAGACGATCT GCACACTGGT ATTTCGGTTT TTGGGGCCGC GGGCGGCGAC GGGGCCCGTG
 2321 CGTCCCAGCG CACATGCATG TTCGGCGAGG CGGGGCCTGC GAGCGCGGCC ACCGAGAATC GGACGGGGGT AGTCTCAAGC
 2401 TGGCCGGCCT GCTCTGGTGC CTGGCCTGCG GCCGCCGTGT ATCGCCCCGC CCTGGGCGGC AAGGCTGGGA GCTCAAAATG
 2481 GAGGACGCGG CGCTCGGGAG AGCGGGCGGG TGAGTCACCC ACACAAAGGA AAAGGGCCTT TCCGTCCTCA GCCGTGCGTT
 2561 CATGTGACTC CACGGAGTAC CGGGCGCGGT CCAGGCACCT CGATTAGTTC TCGAGCTTTT GGAGTACGTC GTCTTTAGGT
 2641 TGGGGGGAGG GGTTTTATGC GATGGAGTTT CCCCACACTG AGTGGGTGGA GACTGAAGTT AGGCCAGCTT GGCAC TTGAT
 2721 GTAATTCTCC TTGGAATTTG CCTTTTGTGA GTTTGGATCT TGGTTCATTC TCAAGCCTCA GACAGTGGTT CAAATTAATA
 2801 CGACTCACTA TAGGGAGACT TCTTTCTCCC ATTTCAAGTG TCGTAAGCTA TCAAACAAGT TTGTACAAAA AAGCAGGCTT
 2881 CGCCACCATG ACTTCACCAA ACGAACTAAA TAAGCTGCCA TGGACCAATC CTGGAGAAAC AGAGATATGT GACCTTTCAG
 2961 ACACAGAATT CAAATATCT GTGTTGAAGA ACCTCAAGGA AATTCAAGAT AACACAGAGA AGGAATCCAG AATTCTATCA
 3041 GACAAATATA AGAAACAGAT TGAAATAATT AAAGGGAATC AAGCAGAAAT TCTGGAGTTG AGAAATGCAG ATGGCACACT
 3121 TCACCATCAC CATCACCAIT GAAACCCAGC TTTCTGTAC AAAGTGGTTC GATGGCCGCA GGTAAGCCAG CCCAGGCTC
 3201 GCCCTCCAGC TCAAGGCGGG ACAGGTGCCC TAGAGTAGCC TGCATCCAGG GACAGGCCCC AGCCGGGTGC TGACACGTCC
 3281 ACCTCCATCT CTCTCTCAGG TCTGCCCGGG TGGCATCCCT GTGACCCCTC CCCAGTGCCT CTCCTGGTCG TGAAGGTGC
 3361 TACTCCAGTG CCCACCAGCC TTGTCCTAAT AAAATTAAGT TGCATCATTT TGTTTGACTA GGTGTCCTTG TATAATATTA
 3441 TGGGGTGGAG GCGGGTGGTA TGGAGCAAGG GGCCCAAGTT AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA
 3521 ATAGCATCAC AAATTTTACA AATAAAGCAT TTTTITCACT GCATTCTAGT TGTGGTTTGT CCAAATCAT CAATGTATCT
 3601 TATCATGTCT GGATCCGCTT CAGGCACCGG GCTTGCGGGT CATGCACCAG GTGCGCGGTC CTTGCGGCAC CTCGACGTCTG
 3681 GCGGTGACGG TGAAGCCGAG CCGCTCGTAG AAGGGGAGGT TGCGGGGCGC GGAGGTCTCC AGGAAGGCGG GCACCCCGGC
 3761 GCGCTCGGCC GCCTCCACTC CGGGGAGCAC GACGGCGCTG CCCAGACCCT TGCCCTGGTG GTCGGGCGAG ACGCCGACGG
 3841 TGGCCAGGAA CCACGCGGGC TCCTTGGGCC GGTGCGGCGC CAGGAGGCTT TCCATCTGTT GCTGCGCGGC CAGCCTGGAA
 3921 CCGCTCAACT CGGCCATGCG CGGGCCGATC TCGGCGAACA CCGCCCCCGC TTCGACGCTC TCCGGCGTGG TCCAGACCGC
 4001 CACCGCGGCG CCGTCGTCCG CGACCCACAC CTTGCCGATG TCGAGCCCGA CGCGCGTGAG GAAGAGTTCT TGCAGCTCGG
 4081 TGACCCGCTC GATGTGGCGG TCCGGGTGGA CGGTGTGGCG CGTGGCGGGG TAGTCGGCGA ACGCGGCGGC GAGGGTGCCT
 4161 ACGGCCCGGG GGACGTGCTC GCGGGTGGCG AGGCGCACCG TGGGCTTGTA CTCGGTCATG GTGGCTGCA GAGTCGCTCT
 4241 GTGTTGAGG CCACACGCGT CACCTTAATA TGCGAAGTGG ACCTGGGACC GCGCCGCCCC GACTGCATCT GCGTGTTTTC
 4321 GCCAATGACA AGACGCTGGG CGGGGTTTGT GTCATCATAG AACTAAAGAC ATGCAAATAT ATTTCTTCCG GGGACACCGC
 4401 CAGCAAACGC GAGCAACGGG CCACGGGGAT GAAGCAGCTG CGCCACTCCC TGAAGATCCC CCTTATTAAC CCTAAACGGG
 4481 TAGCATATGC TTCCCGGTA GTAGTATATA CTATCCAGAC TAACCCTAAT TCAATAGCAT ATGTIACCCA ACGGGAAGCA
 4561 TATGCTATCG AATTAGGCTT AGTAAAGGG TCCTAAGGAA CAGCGATCTG GATAGCATAT GCTATCCTAA TCTATATCTG
 4641 GGTAGCATAT GCTATCCTAA TCTATATCTG GGTAGCATAG GCTATCCTAA TCTATATCTG GGTAGCATAT GCTATCCTAA
 4721 TCTATATCTG GGTAGTATAT GCTATCCTAA TTTATATCTG GGTAGCATAG GCTATCCTAA TCTATATCTG GGTAGCATAT

4801 GCTATCCTAA TCTATATCTG GGTAGTATAT GCTATCCTAA TCTGTATCCG GGTAGCATAT GCTATCCTCA TGCATATACA
 4881 GTCAGCATAT GATACCCAGT AGTAGAGTGG GAGTGCTATC CTTTGCATAT GCCGCCACCT CCCAAGGAGA TCCGCATGTC
 4961 TGATTGCTCA CCAGGTAAAT GTCGCTAATG TTTTCCAACG CGAGAAGGTG TTGAGCGCGG AGCTGAGTGA CGTGACAACA
 5041 TGGGTATGCC CAATTGCCCC ATGTTGGGAG GACGAAAATG GTGACAAGAC AGATGGCCAG AAATACACCA ACAGCAGCA
 5121 TGATGTCTAC TGGGGATTTA TTCTTTAGTG CGGGGAATA CACGGCTTTT AATACGATTG AGGGCGTCTC CTAACAAGTT
 5201 ACATCACTCC TGCCCTTCCT CACCTCATC TCCATCACCT CCTTCATCTC CGTCATCTCC GTCATCACCC TCCGCGCAG
 5281 CCCCTTCCAC CATAGGTGGA AACCAGGGAG GCAAACTAC TCCATCGTCA AAGCTGCACA CAGTCAACCCT GATATTGCAG
 5361 GTAGGAGCGG GCTTTGTCTAT AACAAAGTCC TTAATCGCAT CCTTCAAAAC CTCAGCAAAT ATATGAGTIT GTAAAAAGAC
 5441 CATGAAATAA CAGACAATGG ACTCCCTTAG CGGGCCAGGT TGTGGGCGG GTCCAGGGGC CATTCCAAAG GGGAGACGAC
 5521 TCAATGGTGT AAGACGACAT TGTGGAATAG CAAGGCGAGT TCCTCGCCTT AGGTTGTAAA GGGAGGTCTT ACTACCTCCA
 5601 TATACGAACA CACCGGCGAC CCAAGTTCCT TCGTGGTAG TCCTTTCTAC GTGACTCCTA GCCAGGAGAG CTCTTAAACC
 5681 TTCTGCAATG TTCTCAAATT TCGGGTTGGA ACCTCCTTGA CCACGATGCT TTCCAAACCA CCTCCTTTT TTGCGCCTGC
 5761 CTCCATCACC CTGACCCCGG GTCCAGTGC TTGGGCCCTC TCCTGGGTCA TCTGCGGGC CCTGCTCTAT CGTCCCGGG
 5841 GGCACGTGAG GCTCACCATC TGGGCCACCT TCTTGGTGGT ATTCAAATA ATCGGCTTCC CCTACAGGGT GGAAAAATGG
 5921 CCTTCTACCT GGAGGGGGCC TGGCGGTGG AGACCGGAT GATGATGACT GACTACTGGG ACTCCTGGGC CTCTTTTCTC
 6001 CACGTCCAG ACCTCTCCCC CTGGCTCTT CACGACTTCC CCCCCTGGCT CTTTCACGT CTCTACCCCG GCGGCCTCCA
 6081 CTACCTCTC GACCCCGGCC TCCACTACCT CCTCGACCC GGCCTCCACT GCCTCTCGA CCCC GGCTC CCGCACCTCC
 6161 TCCAGCCCCA GCACCTCCAC CAGCCCCAGC TCCCCAGCT CCAGCCCCAC CAGCACCAGC CCTCCAGCC CCACCAGCCC
 6241 CAGCCCCCTC GGCACCTCCT CCAGCCCCAG CACCTCCACC AGCCCCAGCT CCCCAGCTC CAGCCCCACC AGCACCAGCC
 6321 CCTCCAGCCC CACCAGCCCC AGCCCCCTCT GTTCCACCGT GGGTCCCTTT GCAGCCAATG CAACTTGGAC GTTTTTGGGG
 6401 TCTCCGACA CCATCTCTAT GTCTTGGCCC TGATCCTGAG CCGCCCGGG CTCCTGGTCT TCCGCTCCT CGTCTCGTC
 6481 CTCTTCCCG TCCTCGTCCA TGGTTATCAC CCCCTCTCT TTGAGGTCCA CTGCCGCCG AGCCTTCTGG TCCAGATGTG
 6561 TCTCCCTTCT CTCCTAGGCC ATTTCCAGGT CCTGTACCTG GCCCCTCGTC AGACATGATT CACACTAAAA GAGATCAATA
 6641 GACATCTTTA TTAGACGAG CTCAGTGAAT ACAGGGAGTG CAGACTCCTG CCCCCTCCA CAGCCCCCCC ACCCTCATCC
 6721 CCTTCATGGT CGCTGTCAGA CAGATCCAGG TCTGAAAATT CCCCATCCTC CGAACCATCC TCGTCCTCAT CACCAATTAC
 6801 TCGCAGCCCG GAAAACTCCC GCTGAACATC CTCAAGATT GCCTCCTGAG CCTCAAGCCA GGCCTCAAAT TCCTCGTCCC
 6881 CCTTTTGGCT GGACGGTAGG GATGGGATT CTCGGACCC CTCCTCTTCC TCTTCAAGGT CACCAGACAG AGATGCTACT
 6961 GGGCAACGG AAGAAAAGCT GGGTGCGGCC TGTGAAGCTA AGATCTGTG ACATCGATGG GCGCGGGTGT AACTCCGCC
 7041 CATCCCGCCC CTAACCTCCG CCAGTTCCGC CCATTCTCG CCTCATGGCT GACTAATTTT TTTTATTAT GCAGAGCCG
 7121 AGGCGCCTC GGCCTCTGAG CTATTCCAGA AGTAGTGAG AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA AAAGCTAATT
 7201 C

Figure 14

BLASTP v NCBI nr

Query= INSP037.pep
(78 letters)

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF
1,446,218 sequences; 465,230,387 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
ref XP_211857.1 hypothetical protein XP_211857 [Homo sapiens]	109	8e-24
ref XP_112161.2 similar to putative RNA binding protein 1 [Ratt...	38	0.041
ref XP_220945.1 similar to keratin 21, type I, cytoskeletal - r...	37	0.069
ref NP_775151.1 cytokeratin 21 [Rattus norvegicus] >gi 125089 s...	37	0.069
gb AAD49229.2 AF159462_1 EHEC factor for adherence [Escherichia ...	35	0.26
gb AAL57562.1 AF453441_46 Efa1 [Escherichia coli]	35	0.26
emb CAB55629.1 lymphostatin [Escherichia coli]	35	0.26
emb CAC81883.1 Efa1-LifA-Tox protein [Escherichia coli]	35	0.26
gb AAA39399.1 ORF1	35	0.34
pir T36223 hypothetical protein SCE39.13c - Streptomyces coelic...	34	0.59

>ref|XP_211857.1| hypothetical protein XP_211857 [Homo sapiens]
Length = 113

Score = 109 bits (273), Expect = 8e-24
Identities = 54/74 (72%), Positives = 63/74 (84%)

Query: 1 MTSPNELNKLPTNPGETEICDLSDTEFKISVLKNLKEIQDNTEKESRILSDKYKKQIEI 60
MTSPNELN+ P TNP ETEIC++ D EFKI+VL+ L EIQDNTEKE ++LSDK K+IEI
Sbjct: 1 MTSPNELNEAPGTPAETEICNILDREFKIAVLRKLNEIQDNTEKELKVLSDKIIKEIEI 60

Query: 61 IKGNQAEILELRNA 74
IK NQAEILEL+NA
Sbjct: 61 IKMNQAEILELKNA 74

Figure 15

BLAST v month-aa

Query= INSP037.pep
(78 letters)

Database: NCBI: Rolling month (30 days) of new/revised protein
sequences

37,755 sequences; 14,558,446 total letters

Searching.....done

	Score (bits)	E Value
Sequences producing significant alignments:		
ref XP_141262.1 similar to NAG14 protein [Homo sapiens] [Mus mu...	30	0.27
ref NP_831679.1 Phage-related protein [Bacteriophage phBC6A51] ...	30	0.36
ref NP_083191.1 RIKEN cDNA 1200008A14 [Mus musculus] >gi 128359...	29	0.61
ref NP_852012.1 neck appendage [Streptococcus phage C1] >gi 309...	28	0.80
ref NP_064648.1 neurexin I; neurexin I beta; neurexin I alpha; ...	28	1.0
ref XP_319358.1 ENSANGP00000006161 [Anopheles gambiae] >gi 2130...	28	1.0
ref XP_308412.1 ENSANGP00000019827 [Anopheles gambiae] >gi 2129...	28	1.0
ref NP_196806.2 expressed protein [Arabidopsis thaliana]	27	1.8
gb AAL29689.1 Snf2-related chromatin remodeling factor SRCAP [T...	27	1.8
ref XP_314825.1 ENSANGP00000011098 [Anopheles gambiae] >gi 2129...	27	1.8
ref XP_311503.1 ENSANGP00000013657 [Anopheles gambiae] >gi 2129...	27	2.3

>ref|XP_141262.1| similar to NAG14 protein [Homo sapiens] [Mus musculus]
 ref|XP_230311.1| similar to NAG14 protein [Homo sapiens] [Rattus norvegicus]
 ref|NP_848840.1| RIKEN cDNA 6430556C10 gene [Mus musculus]
 dbj|BAC28656.1| unnamed protein product [Mus musculus]
 dbj|BAC33302.1| unnamed protein product [Mus musculus]
 Length = 640

Score = 30.0 bits (66), Expect = 0.27

Identities = 22/59 (37%), Positives = 33/59 (55%), Gaps = 8/59 (13%)

Query: 20 ICDLSDTEFK-ISVLKNLKEIQDNTESRILSDKYKKQIEIIKGN-----QAEILEL 71
 +C S+ K I V KNL+E+ D +R+L + ++ QI+IIK N EIL+L
 Sbjct: 50 VCSCSNQFSKVICVRKNLREVDPGISTNTRLL-NLHENQIQIIVNSFKHLRHLEILQL 107

Figure 16A

TBLASTN v NCBI nt-month

Query= INSP037.pep
(78 letters)

Database: NCBI: Rolling month (30 days) of new/revised nt sequences
(GenBank+ EMBL + DDBJ sequences (but no EST, STS, GSS, or phase 0, 1
or 2 HTGS sequences))

44,426 sequences; 216,324,491 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
gb AC093724.3 Homo sapiens BAC clone RP11-1L5 from 2, complete ...	105	2e-23
emb BX510371.4 Human DNA sequence from clone RP13-728A10 on chr...	89	2e-18
gb AC144561.8 Homo sapiens 3 BAC RP11-628C23 (Roswell Park Canc...	82	4e-16
dbj AP001827.5 Homo sapiens genomic DNA, chromosome 11 clone:RP...	80	1e-15
emb Z97632.3 HS196E23 Human DNA sequence from clone RP1-196E23 o...	66	3e-11
emb BX322234.7 Human DNA sequence from clone XXYac-65C7_A on ch...	62	5e-10
dbj AP005138.3 Homo sapiens genomic DNA, chromosome 18 clone:RP...	54	1e-07
dbj AP006292.2 Homo sapiens genomic DNA, chromosome 9 clone:RP1...	54	1e-07
gb AC083903.10 Homo sapiens chromosome UNK clone RP11-785G23, c...	47	1e-05
gb AY293855.1 Homo sapiens insulin-like growth factor 2 recepto...	45	7e-05

>gb|AC093724.3| Homo sapiens BAC clone RP11-1L5 from 2, complete sequence
Length = 161617

Score = 105 bits (263), Expect = 2e-23
Identities = 55/78 (70%), Positives = 62/78 (78%)
Frame = -3

Query: 1 MTSPNELNKLPTWNPGETEICDLSDFKISVLKLNKEIQDNTESRILSDKYKKQIEI 60
MTSPNELNK P NP ET++CDLS EFKI+VL+ LKEIQDNTES RILSDK+ K IEI
Sbjct: 22538 MTSPNELNKAPRINPQETKLCDSLHGEFKI AVLRLKLEIQDNTESGFRILSDKFNKDIEI 22359

Query: 61 IKGNQAEILELRNADGTL 78
I +AEILEL+NA G L
Sbjct: 22358 IFKTRAEILELKNAIGIL 22305

Score = 30.0 bits (66), Expect = 1.7
Identities = 19/60 (31%), Positives = 35/60 (57%)
Frame = +3

Query: 14 NPGETEICDLSDFKISVLKLNKEIQDNTESRILSDKYKKQIEI IKGNQAEILELRN 73
+P + EI DLS+ EFK+ V+K ++E + E + + K +K I+ +KG + ++ N
Sbjct: 111237 DPNKEEITDLSEKFKL-VIKLIREGPEKGEAQCK----KIQKVIQ*VKGETFKEIDSLN 111401

Figure 16B

TBLASTN v NCBI nt

Query= INSP037.pep
(78 letters)

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,
or phase 0, 1 or 2 HTGS sequences)
1,794,754 sequences; 8,367,844,792 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
gb AC112641.3 Homo sapiens 3 BAC RP11-431I8 (Roswell Park Cance...	158	2e-37
gb AC026118.17 Homo sapiens 3 BAC RP11-67F24 (Roswell Park Canc...	158	2e-37
emb AL020989.2 HS192P9 Human DNA sequence from clone RP1-192P9 o...	117	3e-25
gb AC009811.14 Homo sapiens chromosome 3, clone RP11-491K7, com...	116	7e-25
gb AC108166.5 Homo sapiens BAC clone RP11-724L20 from 4, comple...	115	9e-25
gb AC011299.3 AC011299 Homo sapiens BAC clone RP11-232C20 from 7...	115	1e-24
gb AC144613.1 Pan troglodytes chromosome 7 clone RP43-1F6, comp...	115	1e-24
dbj AP001992.4 Homo sapiens genomic DNA, chromosome 11q clone:R...	115	1e-24
emb AL359393.9 Human DNA sequence from clone RP11-338I3 on chro...	114	2e-24
emb AL353577.22 Human DNA sequence from clone RP11-661K19 on ch...	114	2e-24

>gb|AC112641.3| Homo sapiens 3 BAC RP11-431I8 (Roswell Park Cancer Institute Human BAC
Library) complete sequence
Length = 165619

Score = 158 bits (399), Expect = 2e-37
Identities = 78/78 (100%), Positives = 78/78 (100%)
Frame = +3

Query: 1 MTSPNELNKL PWTNPGETEICDLSDFKISVLKNLKEIQDNTEKESRILSDKYKKQIEI 60
MTSPNELNKL PWTNPGETEICDLSDFKISVLKNLKEIQDNTEKESRILSDKYKKQIEI
Sbjct: 47052 MTSPNELNKL PWTNPGETEICDLSDFKISVLKNLKEIQDNTEKESRILSDKYKKQIEI 47231

Query: 61 IKGNQAEILELRNADGTL 78
IKGNQAEILELRNADGTL
Sbjct: 47232 IKGNQAEILELRNADGTL 47285

Figure 17

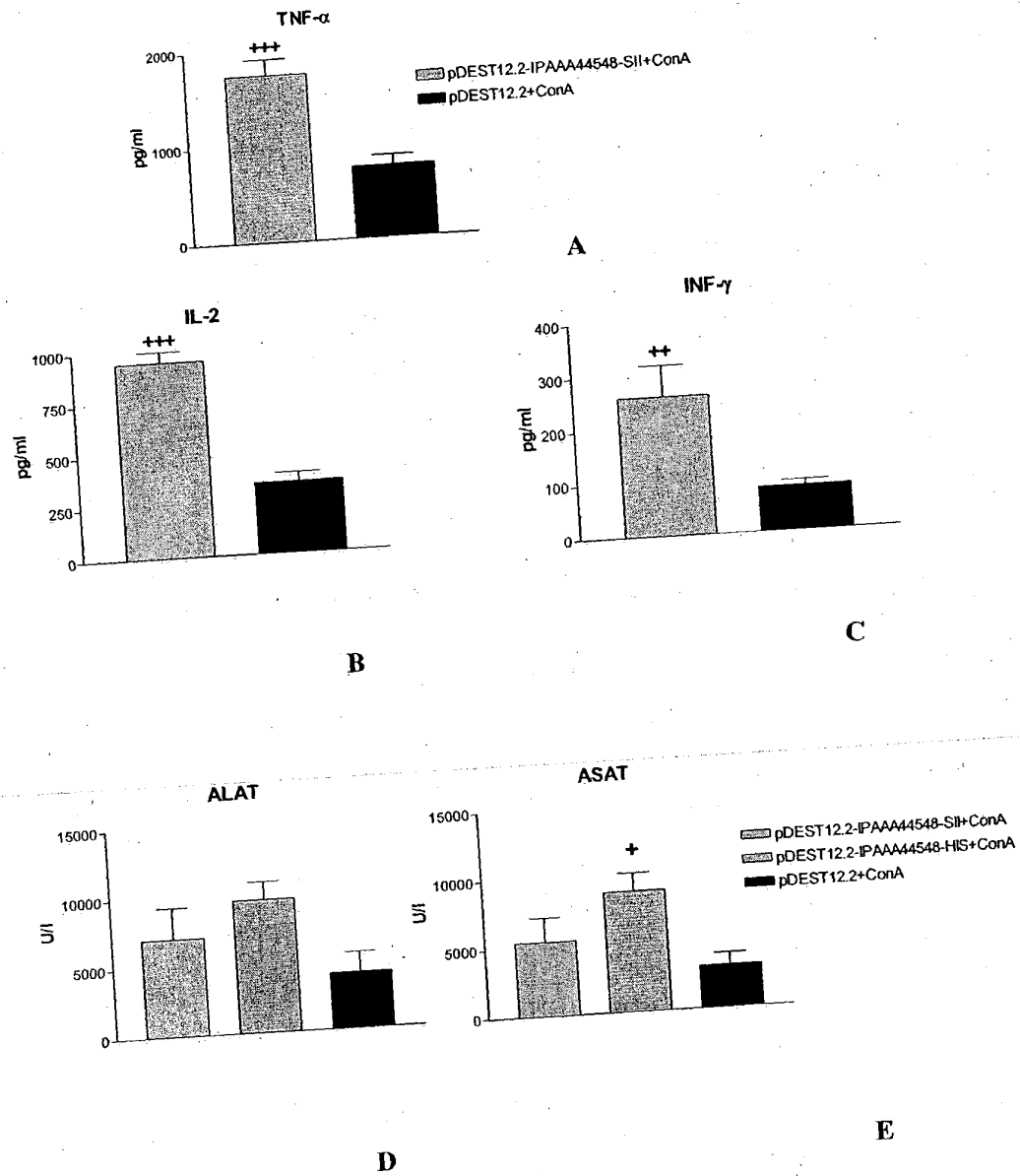


Figure 18

